

## **Historic, Archive Document**

Do not assume content reflects current scientific knowledge, policies, or practices.



Reserve

asB123

.3

.G47

1984

AD-33 Bookplate  
(1-63)

**NATIONAL**

**A  
G  
R  
I  
C  
U  
L  
T  
U  
R  
A  
L**



**LIBRARY**

GERMPLASM RESOURCES INFORMATION NETWORK

---

UNITED STATES DEPARTMENT OF AGRICULTURE

AGRICULTURAL RESEARCH SERVICE

BELTSVILLE AREA

BELTSVILLE AGRICULTURAL RESEARCH CENTER

PLANT GENETICS AND GERMPLASM INSTITUTE

DATABASE MANAGEMENT UNIT

Third Revision

October 5, 1984

## WHAT IS GRIN?

Plant germplasm within the United States is managed by the National Plant Germplasm System (NPGS). One of the major structural components of this system is the Germplasm Resources Information Network (GRIN) which is used for management and operation as well as enhanced communication to scientists regarding the location and characteristics of germplasm they may wish to obtain for research purposes. The major purpose of the GRIN is to serve as a central repository of information concerning major aspects of plant germplasm in the NPGS and to provide ready accessibility of this information to all users of this system.

## WHY HAVE GRIN

### Importance of Germplasm

Plant germplasm is the raw material required by plant breeders for the development of new, superior crop varieties that can ensure a stable, plentiful supply of high quality food, feed and fiber. Most of the plants from which the United States derives its food and fiber were introduced from other countries. The list of economically important native plants is very short and includes sunflowers, cranberries, blueberries, strawberries, pecans, hops, range grasses, conifers, and hardwoods.

There are large gaps in the genetic diversity base of some crops, particularly the wild species and primitive varieties that may contain genes for disease and insect resistance and other desirable traits. Although found in many areas of the world, these sources of diversity are rapidly being depleted, displaced, or abandoned. Once lost, these sources will never again be available to mankind. The need for these diverse sources becomes apparent when the genetic vulnerability of present American monoculture is examined while the constant battle against pathogens continues.

### Plant Introductions and the NPGS

The American Government recognized early the need for a continuing search for more adaptable crops. As early as 1819, American consuls

overseas were urged to send useful plants back to the United States. From this start, the essential elements of the present plant germplasm system developed gradually. This system developed into the NPGS whose goal is to provide on a continuing, long term basis the plant genetic diversity needed by farmers and public and private plant scientists to improve productivity of crops and minimize the vulnerability of those crops to biological and environmental stresses.

Minimizing crop losses through control of biological and environmental stresses is far more difficult and costly than through increasing the genetic diversity among varieties of a given crop. Therefore, an NPGS objective is to broaden the genetic diversity of a crop throughout its production areas by having that production come from an array of varieties, all productive but each different from the others in its range of tolerance to one or more potential stresses. Collection and introduction of new germplasm through various means is the first step toward achievement of this goal.

The NPGS now maintains over 400,000 accessions (samples) of germplasm in the form of seed and vegetatively propagated stocks. These accessions are primarily landraces and unimproved germplasm from foreign sources. New accessions are added to the NPGS at a rate of 7,000 to 15,000 per year. The need for the actual accession to enter the germplasm system is paralleled by the need for information about the accession to be available to users of the system. The immense size of the system creates a challenge for information management. Difficulty of obtaining information, lack of uniformity concerning this information, and its overall poor treatment prompted the NPGS to



integrate an information management system as a major structural component. The Germplasm Resources Information Network was designed and developed to act as this system.

## BRIEF HISTORY

A feasibility study was conducted during 1976-77 which investigated and identified the need for information management systems in the efficient collection, conservation, distribution and utilization of plant germplasm in the National Plant Germplasm System. This study drew the following conclusions about the existing information-management system for germplasm. "An information system exists within the plant genetics resources community of the United States but this system lacks the organization, communication techniques, trained personnel, and funding to satisfy the requirement of the NPGS community." From this study the USDA Agricultural Research Service (ARS) recognized the critical need for a nationally unified information system to serve the diverse needs of the NPGS. The NPGS' Germplasm Resources Information Network (GRIN) came about through this realization. The developmental phase of this network was the Germplasm Resources Information Project (GRIP). This project was established under a five-year cooperative agreement with the Laboratory for Information Science in Agriculture (LISA) to develop a computer-based information system.

Analysis of the diverse needs of the germplasm community identified two groups of information users within the NPGS. One group, the suppliers, consists of those who acquire, maintain, and distribute germplasm and data, such as curators and staffs of the National Seed Storage Laboratory (NSSL) and Regional Plant Introduction Stations (RPIS). A second, or demand group, consists of plant breeders, scientists and other researchers who use the germplasm resource and its data.

Further analysis identified specific needs of both groups. Small scale prototypes were then constructed to meet the needs of the supply side as well as to verify the objectives of each site. From the evaluation of these prototypes, a user-oriented approach was selected for design development.

In the evaluation and design of the information system, a centralized computer center was selected to optimize operational speed for all users. A 'Database' concept was adopted to enhance information by reducing redundancy and by 'linking' together in one location most of the pertinent information about a particular germplasm sample (accession) --from its native habitat to the most recent characteristic and evaluation results. This centralization also allows researchers access to a more extensive collection of samples from which to choose, thereby reducing the possibility of overlooking a potentially valuable collection. Maintenance of information is supported through automatic updates that are quickly available to everyone. The most accurate and current information is thus accessible without time consuming notifications. Multilevel system updates are made by individuals and organizations recognized as experts in that particular area. For instance, plant taxonomists will monitor and maintain taxonomic nomenclature, PIO will maintain information concerning the sample's origin and particulars about its introduction, the RPIS' maintain viable samples and serve as points of contact for sample availability, and the breeders, growers and researchers provide evaluation and characteristic information.

To make the data understandable at a National level, Crop Advisory

Committees (CAC) were developed simultaneously with GRIP. These committees, composed of crop experts from all parts of the NPGS, develop standard evaluation and characterization criteria for their particular crop as well as descriptor lists and standard methods of measurement and reporting.

The now completed design phase brought the transformation of GRIP to GRIN. On July 1, 1983 GRIN was transferred to the Plant Genetics and Germplasm Institute (PGGI) within the Agricultural Research Service, USDA, located in Beltsville, Maryland. The management and final development of the network are controlled by the Database Management Unit (DBMU). Implementation of the system was finalized in February, 1984.

## HOW DOES GRIN WORK?

GRIN has three important functions to fulfill. First, it serves as a central repository for valuable germplasm information that is accessible by the entire germplasm community. Second, it is a means for the CAC's to begin standardization of crop descriptors. Third, it provides a mechanism for each of the RPIS' and other sites to handle daily inventory.

Anyone who can justify a need for accessing the GRIN database can obtain permission to use it. Access to the database can only be gained by submitting a request to the DBMU and having a logon and password assigned for the Prime computer. In addition, a password is also required to access the database.

The database is designed to permit flexibility to the users in storing and retrieving information. A network design is one that allows multiple paths to the data but has linkages that connect all the data together. The GRIN Data Model presents a pictorial view of how segments of the information are related. Appendix I contains a list of all data elements in the database as of the date of this document.

Retrieving information from the database is accomplished by executing a Prime procedure called VISTA. All information stored in the database is accessible, however, access to some data is restricted to the site-owner. NOTE: A user will find it essential to understand some basic principles of a database management system and the diagram of the GRIN DATA MODEL. These topics are thoroughly discussed in the Grin

Data Retrieval User Manual. Examples given in Attachment II illustrate extracting information from the database.

PIO has sole responsibility for maintaining accurate passport data, geographic acquisition, and geographic origin information. The Plant Exploration and Taxonomy Laboratory (PETL) will maintain all plant taxonomy. PIO and PETL will be able to modify information in their respective areas, however any user possesses retrieval access. Public users and participating sites are only permitted to retrieve this information. They cannot modify its contents.

The RPIS' and other participating germplasm sites are the owners of the inventory and characteristic data for their respective site. This means these sites are responsible for maintaining accurate characteristic and inventory information. They also have system procedures available for ensuring data integrity. The public users have permission to retrieve all but inventory data from the database; this can only be accessed by its owner.

As stated earlier, the information residing in the database is owned and maintained by sites within the NPGS while the DBMU acts as the caretaker of the system. The DBMU maintains all application computer software (programs), the database management system (DBMS), liaison with the computer operations (Prime minicomputer), and volume data loading. The DBMU also provides technical assistance to users in preparation of software that is unique for a specific site. Database access and system security are also important system management tasks.

Contact the following office for any additional information:

Database Manager

United States Department of Agriculture

Agricultural Research Service

Plant Genetics and Germplasm Institute

Beltsville Agriculture Research Center-West

Building 001, Room 130

Beltsville, Md. 20705

Phone: 301-344-3318

FTS: 8-344-3318

## GRIN HARDWARE

## THE PRIME -

GRIN is contained on a PRIME 750 mini-computer. The 750 is housed, operated and maintained by the Communications and Data Services Division (CDSO) of ARS, located in the National Agricultural Library (NAL), Beltsville Agricultural Research Center (BARC) in Beltsville, Maryland.

The 750 has a 32-bit architecture and currently has the following hardware features:

- 3.5 Million bytes (Mb) of main memory

- 2400 Mb of user available disk space (1500 Mb owned by GRIN)

- 8 300 Mb disk drives, 2 controllers

- 1 9-Track 1600/6250 bpi Tape Drive

- 3 Synchronous Communications Ports (1 9600 and 2 4800 Baud)

- 48 Asynchronous Communications Ports:

- 15 Direct 9600 Baud

- 8 Dialups 1200 Baud

- 4 Dialups 300 Baud

- 12 Telenet 300/1200 Baud

- 9 currently unused

Prime software features include: DBMS; F77 (Fortran ANSI X3.9-1978); DPTX (3270) Communication with the IBM mainframe at the Washington Computer Center; the FORMS package for data entry and manipulation



template creation; INFO8S a relational data management utility; and INFOTEXT; SCREEN and TEXTPLUS, word processing packages; COBOL, FTN (FORTRAN IV), BASIC, BASICVM and PL1G compilers. Through a NETLINK to another Prime 750 also located at NAL, users have full HASP support, as well as electronic mail.

#### Terminals -

GRIN software written for fully participating curatory sites and inventory managers uses the FORMS package and assumes the user's terminal is a Perkin- Elmer OWL model 1251.

Many other teminal types may use the special forms capabilities of INFO although these have not been verified by the DBMU. They are: ACT-IV; ACT5A; Adds Consul; Lear Siegler ADM-1, ADM-42 and Xerox 850 Word Processor; Lear Siegler ADM-3A and ADM-31; Anderson Jacobson model 510/510A; ANSI; Adds Regent 20, 25, 40, 60, 100 and 200, also Teleray; AT; Beehive; CYBERNEX; Datapoint 8200; DELTA; Datagraphics 132A and 132B; DMD3; Data media Elite 1521; Hazeltine 1510; Hewlett-Packard 2621; IBM 3101; Infoton 100 and 400; Informer 304; DEC VT100; Lynwood; Newbury; Perkin-Elmer OWL, FOX, BANTAM and the OWL 1200 series; SAMP; Soroc; Televideo 912, 920 and 950; Volker- Craig 404 and 414; DEC VT52; ZENITH Z19 and TAB.

SCREEN supports the following terminal types:

Lear Siegler ADM-31, ADM-42; Beehive; Hazeltine 1510; Hewlett-Packard 2621; Infoton; Adds Regency; 40; Perkin-Elmer OWL

1251; TAB 132/15 and 132/15-G; Televideo 912, 920, 925 and 950;  
Terminals Smith TS-1; Volker-Craig 404; Heath ZENITH Z19.

Only the Beehive PT45, Perkin-Elmer Model 1251 (OWL) and Televideo 950  
can make use of FORMS; these three plus the ADM-42 can interface with  
DPTX. Any terminal with RS232C interfaces can use all other Prime  
features.

Telenet -

TELENET users are expected to have ID's and passwords. There are  
TELENET and WATS numbers available for use.

## Appendix I

There are more than 300 unique data elements in the database. The following list describes the area of the database the information resides in, records within the area, and data elements in each record.

NOTE: Char = Character; # = Numeric; & = Real Variable;

## \*STANDARDS AREA\*

## Family Record

30 char	Family name
40 char	Family authority
4 char	PIO code

## Genus Record

30 char	Genus name
40 char	Genus authority
4 char	PIO code

## Species Record

30 char	Species name
40 char	Species authority
4 char	PIO code

## Family Synonym Record

30 char	Family synonym name
40 char	Family synonym authority
12 char	Date created

## Binomial Synonym Record

30 char	Binomial Genus name
30 char	Binomial Species name
40 char	Binomial Genus authority
40 char	Binomial Species authority
12 char	Date created

## Geographical Acquire Record

26 char Acquisition country  
 20 char Acquisition state  
 4 char PIO code

## Geographical Origin Record

26 char Origin country  
 20 char Origin state  
 4 char PIO code

## \*ACCESSION AREA\*

## Accession Record

10 char Primary identifier  
 30 char Subspecies  
 36 char Variety  
 50 char Cultivar  
 30 char Common name  
 12 char Date received  
 12 char Date released  
 6 char Entry logonid  
 6 char Primary supply site  
 10 char Inventory identification  
 12 char Date entered  
 12 char Date PI assigned  
 12 char Date Taxonomy assigned  
 3 char Attribute flag  
 3 char PIO approved flag  
 3 char PIO Donor-Held flag  
 13 char PIO Crop category  
 18 char PIO Life form  
 16 char PIO Form received  
 20 char PIO Improvement status  
 76 char Pedigree

## Accession Supplemental Record

10 char Supplemental label  
 60 char Supplemental narrative

## Accession Re-identification Record



8 char	High value
4 char	Code count
8 char	Fortran format
12 char	Cobal format
4 char	Definition count
70 char	Definition (occurs definition count times, Max is 5 lines)
9 char	Type
5 char	Editcheck

#### Code Record

16 char	Code value
4 char	Definition count
70 char	Definition (occurs definition count time, Max is 3 lines)

#### \*EVALUATION AREA\*

#### Environment Record

20 char	Identification
12 char	Date planted
12 char	Date harvested

#### Environment Narrative Record

20 char	Narrative label
60 char	Narrative

#### Observation Record

10 char	Accession identification
4 char	# Observation identification
4 char	# 51 unique data elements
8 char	# 7 unique data elements
8 char	& 12 unique data elements
10 char	27 unique data elements
30 char	7 unique data elements
72 char	3 unique data elements

NOTE: The descriptor record contains a description of each item in the observation and where the item is at. (i.e. Flower color is in observation item 051)

## \*COOPERATOR AREA\*

## Cooperator Record

20	char	Last name
20	char	First name
30	char	Organization
30	char	Address line 1
30	char	Address line 2
30	char	Address line 3
20	char	City
20	char	State
10	char	Zip code
26	char	Country
12	char	Telephone number
6	char	Region
8	char	# Identification number
6	char	Update logon (who entered cooperator)
12	char	Update date
6	char	Status
3	char	Class

## Cooperator Group Record

6	char	Site
20	char	Name
6	char	Data Manager
4	char	Access class

## Membership Record

10	char	Member role
----	------	-------------

## Donor Link Record

35	char	Accession identification
70	char	Narrative

## \*SITE AREA\*

## Supply Site Record

6	char	Code
---	------	------

30 char Name  
 30 char Organization  
 30 char Address line 1  
 30 char Address line 2  
 30 char Address line 3  
 20 char City  
 20 char State  
 10 char Zip Code  
 26 char Country  
 12 char Telephone  
 6 char Region  
 20 char Curator first name  
 20 char Curator last name  
 4 char # Last order

#### Generic Order Record

6 char Site identification  
 4 char Order number  
 2 char Type  
 12 char Date required  
 20 char Last name  
 20 char First name  
 30 char Organization  
 30 char Address line 1  
 30 char Address line 2  
 30 char Address line 3  
 20 char City  
 20 char State  
 10 char Zip Code  
 26 char Country  
 10 char Requestor reference  
 60 char Curator comment  
 4 char # Accession count  
 4 char # Accession shipped  
 4 char # Accession split  
 30 char File name  
 76 char Special Instructions (occurs 5 times)  
 12 char Date entered  
 12 char Date shipped  
 6 char Status  
 3 char File lock

#### Site Crop Record

6 char Site  
 20 char Crop Identification  
 4 char & Hundred seed weight  
 4 char & Conversion quantity  
 4 char & Critical replenishment



4 char & Critical distribution  
 4 char # Critical retest interval  
 4 char # Critical germination  
 2 char Measurement unit  
 4 char & Ship quantity  
 4 char Maintenance technique

#### Inventory Record

6 char Site  
 10 char Accession identification  
 10 char Inventory identification  
 12 char Date received  
 12 char Date released  
 10 char Available comment  
 4 char & Quantity on hand  
 4 char & Quantity available  
 12 char Date planted  
 12 char Date harvested  
 10 char Location  
 4 char Pollination code  
 10 char Parent  
 60 char Comments  
 4 char # Maintenance count  
 4 char & Hundred seed weight  
 4 char & Conversion quantity  
 4 char & Critical replenishment  
 4 char & Critical distribution  
 4 char # Critical retest interval  
 4 char # Germination  
 2 char Measurement unit  
 4 char & Shipping quantity  
 2 char # Germination number  
     Germination (occurs germination number times,  
                     max of 4 groups)  
 4 char # Percent normal seedling  
 4 char # Percent hard seed  
 4 char # Percent germination  
 4 char # Germination year  
 12 char Date entered  
 4 char Maintenance  
 8 char Availability flag

#### Inventory Group Record

6 char Site  
 20 char Name  
 6 char Data manager

Inventory Group Link Record

20 char    Group link narrative

Supplier Link Record

20 char    Supplier link narrative

## Appendix II

The following two examples show some of the typical VISTA procedure commands required to extract information from the database and display it on the computer terminal screen.

Example 1 illustrates a retrieval for beet characteristic data contained in the NC7-BETA dataset, but limited to skin color of white, pink, or red.

Example 2 illustrates a retrieval for chickpea characteristic data contained in the W6-CHICKPEA dataset, but limited to only those observations that have a flower set of 36 or fewer days.

```
OK> vista
[DBMS/QUERY Rev. 19.1.3]
> use subschema NC7-BETA of SCHEMA GRIN1-0
> unlock retrieval of areas evaluation-1-area,evaluation-2-area,~
> dictionary-area with 'XXXXX'
> unlock rest of records dataset-record,descriptor-record,~
> observation-record with 'XXXXX'
> select from observation-record where skin-col = 'WHPIRD' ~
> and dataset-query-name = 'NC7-BETA'
Total number of virtual records: 16
```

```
> display using nc7-beta2
```

## NC-7 Beta characteristic data report 2

ID NUMBER	SDLG LF-SZ1	SDLG LF-SZ1	ROOT LEN1	ROOT LEN2	ROOT DIAM1	ROOT DIAM2	ROOT WGT1	ROOT WGT2
PI 109040	5	7	15	27	6	13	223	1036
PI 116906	7	9	11	35	2	14	64	1463
PI 117115	7	7	14	25	3	8	186	393
PI 120701	3	7	7	16	4	6	46	193
PI 140353	7	7	7	15	4	12	38	819
PI 140356	7	7	7	16	4	9	61	542
PI 140361	5	7	6	11	5	14	82	1121
PI 171505	5	8	9	26	4	9	41	795
PI 172729	4	7	13	20	3	10	54	1007
PI 174063	5	7	7	37	5	15	85	3761
PI 176423	7	7	10	13	6	15	195	2057
PI 177276	7	7	12	21	8	13	275	712
PI 182144	6	6	7	18	6	11	181	710
PI 182146	5	8	14	22	6	10	224	786
PI 205987	6	6	8	28	4	11	57	1120

PI 251042          6            6            6            24            3            13            61            2036

> display using nc7-beta5

NC-7 Beta characteristic data report 5

ID NUMBER	SOURCE	LIFE CYC	SDLG STEM-C	SDLG LEAF-C	MATR FOLG-C	DESIGN TYPE	ROOT SHAPE
PI 109040	TURKY	BI	GNYE		GN	MA	TOLOCO
PI 116906	AFGH	ANBI	RDGN	GN	GN	SGMA	TOLOCO
PI 117115	TURKY	ANBI	RDGN	GN	GN	SGMA	TOLO
PI 120701	TURKY	ANBI	RDGN	GN	GN	MALF	LOCO
PI 140353	IRAN	BI	RDGN	GN	GN	MA	FLTOL
PI 140356	IRAN	ANBI	RDGN	GN	GN	SGMA	FLTOL
PI 140361	IRAN	ANBI	RDGN	GN	GN	MATB	GLTOHL
PI 171505	TURKY	ANBI	RDGN	GN	GN	SGMALF	LOCO
PI 172729	TURKY	BI	RDGN	GN	GN	SGMA	TOLO
PI 174063	TURKY	BI	RDGN	GN	GN	MA	GLTOHLL
PI 176423	TURKY	BI	RDGN	GN	GN	SGMA	TOLO
PI 177276	SYRIA	ANBI	RDGN	GN	GN	SGMA	TO
PI 182144	TURKY	BI	RDGN	GN	GN	MA	TOHLC
PI 182146	TURKY	BI	RDGN	GN	GN	SGMA	TOLO
PI 205987	SWEDN	BI	RDGN	GN	GN	SGMATB	TOLO
PI 251042	YUGO	BI	GN	GN	GN	SGMA	TOLO

> quit

do you wish to terminate this session? y

OK> vista

[DBMS/QUERY Rev. 19.1.3]

> use subschema w6-chickpea of schema grinl-0

> unlock retrieval of areas evaluation-1-area,evaluation-2-area,~

> dictionary-area with 'XXXXX'

> unlock rest of records dataset-record,descriptor-record,~

> observation-record with 'XXXXX'

> select from observation-record where dataset-query-name = 'W6-CHICKPEA' ~

> and flwrset > '0' and flwrset le '36'

Total number of virtual records: 12

> display using public format w6-chickpeal

W-6 Chickpea Characteristic Data Report 1

ID NUMBER	IDENT	DSCDE	STAND	FLWST	FLWCL	STMCL	HABIT
PI 212891	1	1	0	36	4	2	3
PI 212891	2	1	0	36	4	2	3
PI 212892	1	1	0	36	4	0	0

PI	212892	2	1	0	36	4	0	0
PI	222771	1	1	0	36	4	2	3
PI	222771	2	1	0	36	4	2	3
PI	254550	1	1	0	35	5	1	2
PI	254550	2	1	0	35	5	1	2
PI	273880	1	1	0	36	4	2	3
PI	273880	2	1	0	36	4	2	3
PI	360224	1	1	0	35	1	1	1
PI	360224	2	1	0	35	1	1	1

> display using public format w6-chickpea2

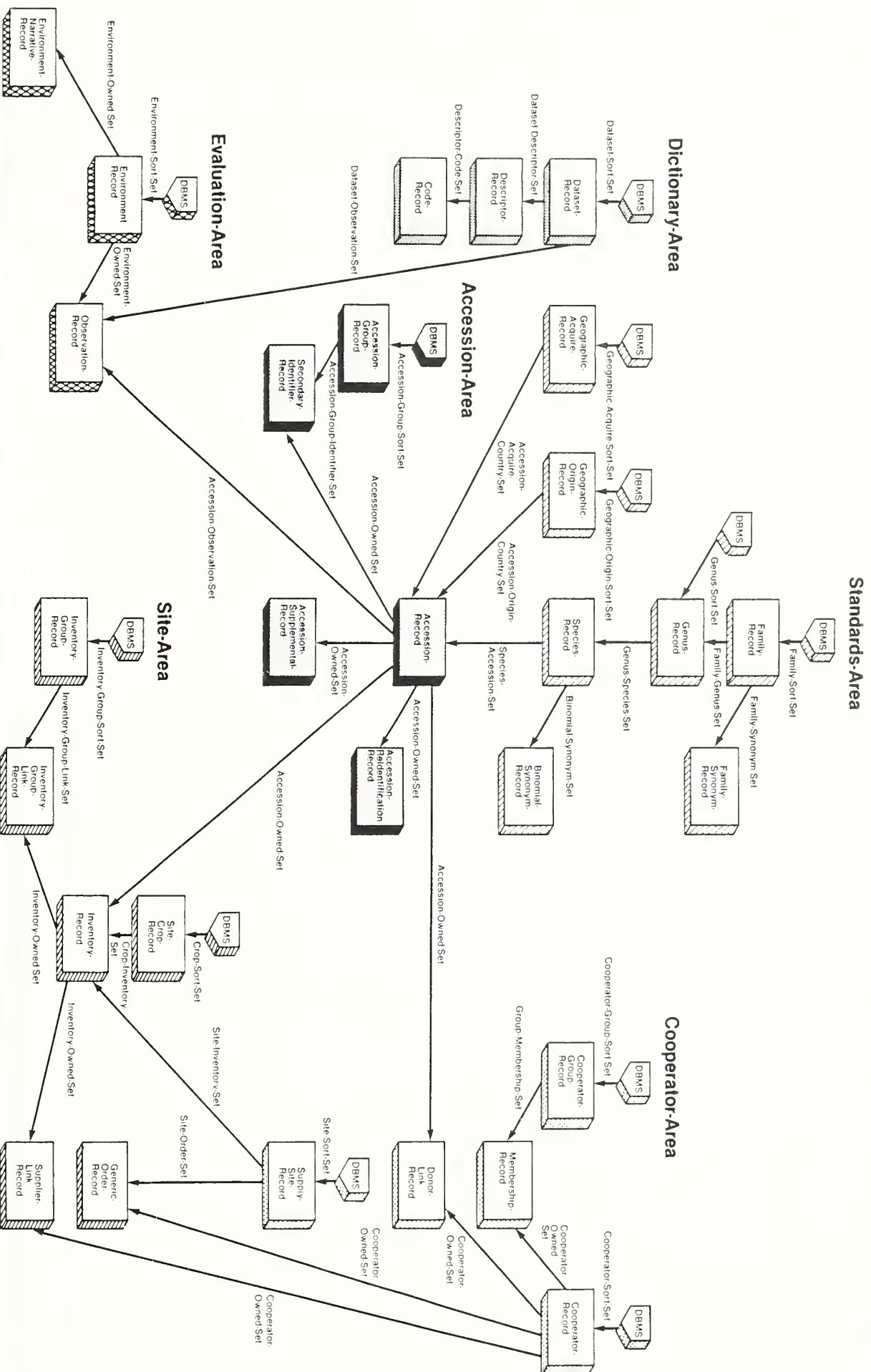
#### W-6 Chickpea Characteristic Data Report 2

	ID NUMBER	RSIZE	SPRED	MATUR
PI	212891	2	3	86
PI	212891	2	3	86
PI	212892	0	0	75
PI	212892	0	0	75
PI	222771	2	3	81
PI	222771	2	3	81
PI	254550	3	4	83
PI	254550	3	4	83
PI	273880	1	3	80
PI	273880	1	3	80
PI	360224	2	3	83
PI	360224	2	3	83

> quit

Do you wish to terminate this session? y









# DATABASE MANAGEMENT UNIT

## GERMPLASM RESOURCES INFORMATION NETWORK (GRIN)

Jim Mowder	Database Manager ADP Technical Approval for Beltsville Area	344-3318
Shirley Freeland	Secretary	344-1666
Erick Abadie	Database Queries & Reports Assist National Small Grain Collection	344-1775
John Belt	Public User Access Discover Prime Hardware Communications	344-2646
Mark Bohning	Crop Advisory Committees	344-1145
Len Jansen and Quinn Sinnott	Scientific Data Management (Data Preparation, Validation and Loading)	344-1775 344-3023
Mark Perry	IBPGR Documentation GRIN Dictionary	344-3133
Rob Selvage	PRIME DBMS Software Database Analysis, Design, Programming Responsible for Maintenance of all Databases	344-3699
Krzysztof Obrebski and Ed Bird and Kurt Endress	Database Design Computer Programming Microcomputer Expertise Graphics	344-3699 344-3095 344-3095



an

